NGST's Scientist's Expert Assistant: Evaluation Results

Anuradha Koratkar [a], Chris Burkhardt [a], Mark Fishman [c], Sandy Grosvenor [d], Jeremy Jones [b], Ray Lucas [a], Lamont Ruley [b], and Karl Wolf [c]

[a]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218
[b]NASA Goddard Space Flight Center, Greenbelt, MD 20771
[c]AppNet Inc., 1100 West Street, Laurel, MD 20707
[d]Booz-Allen Hamilton, 7404 Executive Place, Seabrook, MD 20706

ABSTRACT

This paper describes the approach and evaluation results of the Next Generation Space Telescope (NGST) Scientist's Expert Assistant (SEA) project. The plan describes the goals, and methodology for the evaluation. The objective of this evaluation is to provide a means for the targeted user community to provide feedback to the developers, and to determine if the advanced technologies investigated as part of SEA have achieved the goals that were to be its success criteria.

We can with confidence say that visual, interactive tools in SEA were found to be highly useful by the users. On a scale of 1-5, where 1 was excellent and 5 was poor, the SEA as a whole ranked as 1.7, i.e., between excellent and above average.

Keywords: User support tools, visual tools, proposal development, evaluation of tools, exposure time calculators, target location, target orientation, proposal preparation

1. SCIENTIST'S EXPERT ASSISTANT PROJECT GOALS

In a resource tight environment, an observatory still has to maximize scientific returns. This will be no less true for the Next Generation Space Telescope (NGST). Although NGST is expected to be a much simpler spacecraft than Hubble Space Telescope (HST), its design and environment will have their own peculiarities. Further, the task of obtaining an accurate, feasible set of observations from every proposer will still be critical for smooth operation of the spacecraft. The Scientist's Expert Assistant (SEA) is an experiment to determine if new user support tools using advanced technologies can help achieve this goal (see Jones et. al. 2000 in this volume[2]). In terms of evaluating SEA, the two specific goals that need to be met are:

- Reduction in the need for Contact Scientist/Program Coordinator level of support for routine observations.
- Self-sufficiency of the observer, i.e., observers must find it easier to create accurate, feasible observations.

In designing/developing SEA we considered the present strategy used for user support at HST. Space Telescope Science Institute (STScI) which operates HST, provides HST observers with a suite of software tools as well as expert human support throughout the process – from proposal preparation to data analysis. For SEA we decided to look at the proposal preparation process, which is the front end of the operations up to the point where the proposal is received at STScI. We found that the need for extensive user support at STScI is not only due to the complexity of the spacecraft operations, but also due to evolution in spacecraft operations, instrument/detector sensitivities, etc. As the telescope operations change in time, so does the information which must be communicated to the user community. Further, the programs submitted by the user community have to reflect this evolving information and still produce accurate and feasible commands for the telescope to execute. These last few points are true for all observatories, not only HST. Thus, the SEA experiment set out to determine if new user support tools using advanced technologies can help reduce the need for the extensive human user support for routine observations while still providing the personal care/service/advice that is available at STScI, i.e., "can we do more for less".

Not only is obtaining an accurate and feasible set of observations the primary goal of any proposal preparation tool, this information must be obtained in a manner that is cost effective and user friendly. Hence, SEA also aimed at making the observer self sufficient, i.e., make it easier for observers to create accurate, feasible observations. To achieve this goal, the following facts have been guiding the SEA tool development. It was essential that tools

- Let observers focus on science and not on irrelevant technical manipulation.
- Provide easy access to technical and reference material.

- Are intuitive and easy to use, so that different levels of users can be accommodated.
- Allow visualization and interactive manipulation of the observing program.
- Eliminate continual manual re-entry of information.

Thus, SEA's prototyping effort priorities have been visualization, dynamic interactivity and expert systems.

In particular the SEA evaluation plan will determine

1. Does the user find the interactive and graphical interface of SEA useful?
2. Does the environment provide easy access to technical and reference material?
3. Do observers focus on science or technical manipulation?
4. Does the on-line help assist the user?
5. Do the tools provide relevant scientific assistance to the user?
6. Do the tools accommodate different levels of users?
7. In general, do users perceive the SEA as better than the current HST Phase II proposal creation process and associated tools?

2. SELECTION OF TESTBED

To determine the effectiveness of the SEA ideas we set out prototyping the HST's Advanced Camera for Surveys (ACS) because ACS provided us with a real instrument with a good balance of complexity and operational style when compared to the expectations for NGST. However, the ACS launch was delayed and the SEA project had to be evaluated by March 2000. Hence, halfway through the project we changed our testbed and decided to prototype SEA for HST's Wide Field Planetary Camera 2 (WFPC2).

Changing of the testbed provided us with an opportunity to see if ideas developed for one instrument could be easily extended to a new instrument. We found it relatively easy to do this transition. Yet, there were a few "gotchas". There were simple parameters that scientifically mean something to the user, but are implemented very differently for the two instruments. An example of this is CR-SPLIT, which is implemented in different ways for the ACS (CR-SPLIT has an integral value) and for the WFPC2 (CR-SPLIT has either a value yes or no and when it is yes it can only split the exposure in two). One can easily account for these implementation differences behind the scenes in the tools, and yet leave a common interpretation to the parameter in the tools. This example illustrates the need for an instrument expert to be involved in the development of the tools from the beginning, so that commonalties between instruments and operation modes can be seen up front during the design phase of the tools.

3. SEA EVALUATION METHODOLOGY

3.1 Evaluator Profiles:

The SEA evaluators were chosen from the accepted pool of HST Cycle 9 WFPC2 programs. Our original plan was to have at least 21 evaluators so that both the evaluators and programs would cover a range of user expertise and range of program types (see Table 1), and allows us to answer the many questions in a "semi-statistical" manner. Due to time and budget constraints we choose ~15 principal investigators (PI; ~9 from within the Baltimore Washington area and the remaining 6 from the Los Angeles area) from the Cycle 9 accepted pool for WPFC2. We obtained background on them by distributing a questionnaire to each of the prospective evaluators. Our final choice of evaluators included post-doctoral fellows to senior faculty and spanned the full range of familiarity with HST's Remote Proposal Submission process (RPS2), i.e., very familiar with RPS2 (have used it for more than three proposals) to not familiar (have only seen colleagues use it).

Table 1: Choice of programs to be considered for evaluation

PI expertise level	Type of proposal	Size of proposal
New PI	Simple point and shoot programs	small with 1-2 targets
Moderately experienced	Programs using some special scheduling requirements	large > 10 targets
Expert	One program which would be considered tough for various reasons during the strategy phase	

3.2 Evaluation Elements:

There were two stages to the evaluation. The evaluation plan also included some amount of usability testing with the objective of providing a means for the targeted user community to provide feedback on the SEA to the software developers.

The first stage of the evaluation process was to conduct some test scenarios that would allow limited usability test. This phase was also intended to familiarize an evaluator with the many features of the SEA, so that they can get to the second stage. In the second stage, the evaluators were to construct their Cycle 9 Phase II proposal. Both stages of the proposed evaluation process were expected to take approximately half a day. We soon found out that our user community resisted committing too much of their time for the evaluation process. No more than 2 hours could be asked for up front, although on the day of the evaluation depending on the evaluator we spent up to 5 hours!

We had a well-defined set of tasks for the evaluation, but soon found out that most evaluators did not want to follow that strict path. They wanted to freely move within the SEA tool and try out their favorite observing strategies. Since SEA was designed to accept different proposal preparation strategies, we adapted the evaluation plan for each user. We tried to accommodate each evaluator's scientific interests as we introduced them to the SEA. Since we were adapting the various elements in our evaluation strategy, we took care to make sure that all evaluators were introduced to key functionalities of the SEA, and only then were they asked to complete the evaluation form provided. Also, since the evaluators worked primarily with their favorite objects and observing strategies, we did not request that their Phase II program be developed in SEA. We do not consider this a major setback, since our goal had been to see how much of a Phase II proposal could be generated in SEA. We knew that SEA does not have all the functionality of RPS2, hence, we did not expect more than one or two proposals to actually generate a Cycle 9 RPS2 compatible proposal. SEA evaluations were conducted over the period 7th Feb 2000 – 24th Feb 2000. In this paper we present our preliminary analysis of the data.

The evaluations were conducted at the evaluators home institute so that we could also understand platform and connectivity issues. These evaluators were mostly using the Solaris platform, but we also did some evaluation on the Windows and DEC Alpha platforms. We did not run into any connectivity issues. The speed of installation depended on the memory in the machines. The InstallAnywhere software by ZeroG Software worked well. Prior to our evaluations, we had never tried to run SEA on a DEC Alpha, and Java's true platform independence was tested in-situ. In this case we crashed when the expert-system was invoked, but otherwise SEA performed nominally on the DEC Alpha.

The evaluation could not be done electronically since we wanted to understand any relevant behavior or comments made by the evaluators as they prepared their test proposals. During the evaluation we had three observers: an evaluation coordinator (a astronomer), a developer, and a program coordinator who was familiar with HST user support. The evaluator did all the computer input and was observed and coached by the evaluation team. The coordinator guided/coached the evaluator through all the SEA capabilities as needed but tried to allow the evaluator time to discover the solution on their own without getting frustrated. The program coordinator and developer took notes on the results, process and reactions of both the SEA and evaluator. After completing the online tasks the evaluators were asked to complete a survey, the results of which are in the next section, where they ranked several features of SEA in both quality and value. With the above described evaluation strategy we were able to draw some conclusions about how users construct their observing programs, and the level of expert information that they need.

An important output of our evaluation is that we saw a number of experts use different strategies in how they translate a scientific idea into a proposal. The SEA developers all took part in the evaluation process and unanimously found the experience valuable. In a traditional software development scenario, developers rarely get to spend quality time with users.

We concluded that this kind of interaction between developer and user is better than the more traditional restricted method where developers and users interact primarily via a set of formally approved requirements. Our evaluation gave greater insight into how the end-user intuitively tried to use the tools. Although our evaluation strategy is time consuming and expensive, we strongly feel that we should have small formal evaluations through out the development process. This will save costs and time in the long-run by minimizing re-design and re-testing late in the development process.

4. SEA EVALUATION RESULTS:

4.1 Raw Results:

In this section we have tabulated the preliminary results of our evaluation. As of the writing of this paper, we have not yet received all the completed evaluation forms. The results below are based on 75% of our sample.

The following results were obtained when users were asked to rate the SEA features/functionality on a scale of 1- 5, where 1 means most useful/excellent and 5 means not useful/poor.

Table 2: Evaluation of the Visual Target Tuner Module:

Feature or Functionality	Grade
Ability to display the Digitized Sky Survey	1.1
Ability to display the HST field-of-view on the Digitized Sky Survey image, to see the footprints of all the instrument apertures simultaneously	1.2
Ability to graphically and dynamically manipulate any of the HST instrument apertures	1.3
Ability to display the guide star catalog and then filter the catalog objects dynamically	1.5
Access to the various archives such as NED or SIMBAD to obtain target coordinates and other relevant information regarding the field-of view	1.6
Ability to capture scientific constraints and then optimize a feature, e.g., the include/exclude feature that can be used to determine the range of available orientation angles	1.8
Availability of image options such as the color tables etc.	2.0
Ability to access the Digitized Sky Survey via a batch process, so that multiple images can be retrieved in one run	2.6

Table 3: Evaluation of the Exposure Time Calculator Module:

Feature or Functionality	Grade
Graphical access to many different spectral energy distributions	1.4
Ability to access and manipulate multiple parameters (target/instrument) simultaneously	1.5
Graphical access to operational instrument response curves, and ability to compare these with the object's spectral energy distribution	1.6
Ability to compare two sets of exposure parameters	1.8
Ability to compare multiple instrument parameters simultaneously	2.1
Ability to manipulate the graphics display parameters	2.5

Table 4: Evaluation of the Orbit Planner and Visit Planner Modules:

Feature or Functionality	Grade
Graphical display showing detailed overhead information etc. of all the exposures in a visit. Graphical display of all visits in a proposal.	1.2
Ability to graphical and dynamically manipulate exposures in a visit to optimize the scientific returns from that visit.	1.6
Ability to place constraints on the exposures and visits in a more "natural language" way and then to see which exposure/visits are connected to each other with constraints	1.6

Table 5: Evaluation of the General Features in the SEA:

Feature or Functionality	Grade
Instantaneous updates to proposal parameters where possible	1.5
Graphical interface where possible	1.6
Random access to any part of the proposal, i.e., no particular order to developing a proposal	1.6
Ability to develop a proposal interactively, or directly via tables using the multiple table views	1.7
Access to handbooks/reference material	2.2

4.2 Highlights of the evaluation:

Going back to the original questions that were asked in section 1 we can answer them as follows:

- We can with confidence say that visual, interactive tools were found to be highly useful by the users for exploring the multi parameter space. It allowed them the ability to visualize their scientific needs better and they did not have to concentrate on observatory dependent technical aspects of observing.
- On the whole our evaluators felt that preparing proposals using tools that were prototyped in the SEA was a far better experience than the currently available HST proposal preparations tools such as RPS2.
- We anticipated that our better-developed, maturer modules would get a better reception, but we were surprised at the high correlation between the maturity of a module, and how important the evaluators ranked the functionality.
- We found that users perceive certain concepts very differently. For example, users appreciated and liked the context sensitive help and access to reference information, but when we asked them the question "Would you like documentation to be integrated with software" they ranked it as a very low priority.
- By design, SEA does not force a user to develop an observing program in any particular order. Some users quickly adapted and valued this flexibility; others were frustrated by the apparent lack of focus and direction provided by the tool.
- Our attempts at providing expert help were appreciated, but it was felt that more work needed to be done before these ideas were ready for operational use, again indicating that the ranking depended on the maturity of the module[1].
- The instantaneous updating of information was seen as a two edged sword – there were occasions when this was considered essential (i.e. in the VTT), while in other instances it was found to be unacceptable (i.e. in the ETC where users often want to change several parameters at a time and then see the impact).
- Speed of any software has always been an issue with users, we have tried to keep this in mind. During the evaluations we found that this issue was never raised. We feel that this is because of the availability of interactive features and instantaneous updates.
- For much of the reference information, the SEA connects directly to servers at STScI and is highly dependent on internet connectivity. We need to develop strategies to pull all the necessary and relevant information at the start of a session so that a user does not require network connectivity. This is important given that the astronomical community spans across the world and is often travelling.
- A few of our evaluators also saw how many of these tools could be easily extended for other telescopes. They even made suggestions that we should provide users with the ability to ingest information from other telescopes into the SEA.

A greater amount of detail pertaining to the evaluation and the results can also be obtained from our website http://aaaprod.gsfc.nasa.gov/SEA or by contacting our team.

5. CONCLUSIONS

On a scale of 1- 5, where 1 was excellent and 5 was poor, the SEA as a whole ranked as 1.7 (i.e., between excellent and above average). The SEA was an experiment to determine if new user support tools using advanced technologies can help achieve the two specific goals:

- Reduction in the need for Contact Scientist/Program Coordinator level of support for routine observations.
- Self-sufficiency of the observer, i.e., make it easier for observers to create accurate, feasible observations.

We cannot at the present time determine the answer to the first goal. But with regards to the second goal we can say that users found using SEA a better and easier experience than RPS2. In terms of creating accurate and feasible observations we feel that the SEA provided the proof of concept. It is our hope that reduction in support staff will be realized once these capabilities are fully developed and are available for all observers.

6. ACKNOWLEDGEMENTS:

This work was supported by the Next Generation Space Telescope (NGST) technology funding. The SEA is a joint effort between the Space Telescope Science Institute (STScI) and the Advanced Architectures and Automation Branch of Goddard Space Flight Center's (GSFC) Information Systems Center.

7. REFERENCES

1. K. Wolf, et al, "Expert System Technology in Observing Tools", *Proceedings of SPIE Vol. 4010*, 2000.
2. J. Jones et al, "Lessons Learned from the Scientist's Expert Assistant Project", *Proceedings of SPIE Vol. 4010*, 2000.